

## PROCEDE ET DISPOSITIF DE FILTRAGE DE CONTENUS EN LIGNE.

La présente invention concerne un procédé et un dispositif de filtrage de contenus en ligne. Elle vise, en particulier, à protéger les jeunes internautes contre l'accès volontaire ou involontaire à des sites qui ne leur sont pas destinés (contenus à caractère sensible :  
5 pornographie, violence, incitation à la haine raciale).

Les filtres existants, qui sont généralement basés sur le filtrage d'adresses électroniques ("URL" pour Uniform Resource Locator), sont des logiciels qui comparent l'adresse de la page à laquelle l'utilisateur tente d'accéder avec des adresses contenues  
10 dans une base de données. Ces logiciels peuvent être désactivés, comme tout autre logiciel, et l'étendue de leur filtrage est incomplète : leur taux de filtrage atteint, en moyenne, 90 %, c'est-à-dire qu'une page "interdite" sur dix atteint le jeune internaute, ce qui pose un réel problème en milieu scolaire. De plus, l'heuristic des bases de données fait face à une croissance exponentielle des pages web publiées chaque mois, alors que le nombre de  
15 pages indexées mensuellement croît linéairement. La conséquence de ce fait est que de plus en plus de pages web passent et vont passer à côté de l'indexage des solutions fondées sur les bases de données. Les filtres basés sur l'analyse de la couleur "chair" ont également leurs limites et, par un filtrage excessif, interdisent l'accès à toute page où figure la photo d'une personne, par exemple des pages d'information médicale.

20 La présente invention vise à remédier à ces inconvénients.

A cet effet, la présente invention consiste, d'une part, à prévoir un équipement, boîtier externe ou carte interne à l'ordinateur, qui s'intercale entre l'ordinateur (en anglais Personal Computer ou "PC") et le réseau Internet et, d'autre part, à ce que cet équipement mette en  
25 oeuvre un ensemble de règles de décision traitant non seulement le contenu de chaque page mais aussi son environnement (par exemple les pages vers lesquelles mènent les liens disposés sur la page demandée ou les informations de structure, programmatiques ou statistiques, de la page demandée).

Le filtrage peut ainsi filtrer le contenu d'une page dès qu'elle devient accessible et donc de toutes les pages accessibles en ligne, indépendamment de toute base de données  
30 d'URL.

Selon un premier aspect, la présente invention vise un procédé de filtrage de contenu en ligne, caractérisé en ce qu'il comporte :

- de mettre en oeuvre un équipement, boîtier externe ou une carte interne à l'ordinateur, qui s'intercale entre l'ordinateur et un réseau informatique donnant accès à des contenus en  
35 ligne, ledit équipement recevant le contenu provenant du réseau ;
- une étape d'analyse dudit contenu ;
- une étape de recherche d'environnement dudit contenu sur ledit réseau ;

- une étape d'analyse dudit environnement ;
- une étape de décision de filtrage en fonction d'un ensemble de règles de décision dépendante des résultats des étapes d'analyse dudit contenu et de son environnement et
- une étape de transmission ou non dudit contenu audit ordinateur, en fonction du résultat de l'étape de décision de filtrage.

Grâce à ces dispositions, le fonctionnement du boîtier effectue un filtrage non seulement en fonction du contenu auquel l'utilisateur pourrait accéder mais aussi en fonction de l'environnement dudit contenu. De plus, puisque le filtrage est effectué par un boîtier externe, il est moins facile de modifier son fonctionnement qu'un logiciel de filtrage mis en oeuvre sur l'ordinateur. De plus, un équipement autonome peut utiliser ses propres ressources (traitement et/ou mémoire) sans consommer celles de l'ordinateur.

Selon des caractéristiques particulières, au cours de l'étape d'analyse dudit environnement, on traite les pages vers lesquelles mènent les liens hypertextes dudit contenu.

Grâce à ces dispositions, le filtrage est plus fin que lorsque l'on ne traite que le contenu de la page à laquelle l'utilisateur tente d'accéder.

Selon des caractéristiques particulières, au moins une étape d'analyse dudit contenu comporte une première étape de tri rapide de contenu, l'étape de décision comportant une première étape de détermination de décision en fonction du résultat de ladite première étape de tri rapide et, en cas d'indétermination du résultat de ladite première étape de détermination de décision, l'étape d'analyse comporte une deuxième étape de tri de contenu d'une durée supérieure à la première étape de tri rapide, l'étape de décision comportant alors une deuxième étape de détermination de décision en fonction du résultat de la deuxième étape de tri.

Selon des caractéristiques particulières, la première étape de tri rapide de contenu traite un contenu ne comportant pas d'image et en ce que la deuxième étape de tri de contenu comporte une étape de traitement d'image.

Grâce à chacune de ces dispositions, le tri peut être très rapide pour un grand nombre de pages ou contenus accessibles, parce que dès qu'une règle de décision permet la décision, celle-ci est prise. Le tri est néanmoins très précis puisqu'une succession de règles de décision est appliquée, par exemple grâce au traitement d'images et à la compréhension du contenu des images, pour les cas plus complexes.

Selon des caractéristiques particulières, au moins une étape d'analyse comporte une l'étape de traitement d'image au cours de laquelle, pour au moins une image, on analyse la texture du contenu de l'image pour extraire les parties de l'image dont la texture correspond à de la chair.

Grâce à ces dispositions, la détection d'images de chair est plus sûre qu'avec une recherche de la seule couleur chair et la partie visible d'un corps humain représentée par une image peut être déterminée.

Selon des caractéristiques particulières, l'étape de traitement d'image comporte une  
5 étape d'analyse de posture du ou des personnages dont on voit des parties charnelles.

Grâce à ces dispositions, l'analyse du contenu de l'image permet une analyse et une décision de filtrage plus sûre.

Selon des caractéristiques particulières, au moins une étape d'analyse comporte une  
10 étape d'extraction de caractères d'images incorporées dans le contenu en ligne.

Grâce à ces dispositions, les messages textuels présents dans les images peuvent  
être traités pour affiner la compréhension sémantique du contenu en ligne.

Selon des caractéristiques particulières, le procédé tel que succinctement exposé ci-  
dessus comporte une étape d'identification biométrique de l'utilisateur et une étape de  
désactivation de filtrage et d'autorisation d'accès à tout contenu accessible sur le réseau  
15 informatique, en fonction du résultat de ladite identification.

Grâce à ces dispositions, un utilisateur autorisé, par exemple majeur, peut accéder à  
tout contenu accessible en ligne et l'identification de cet utilisateur est plus sûre qu'avec un  
mot de passe et moins contraignante pour l'utilisateur.

Selon des caractéristiques particulières, le procédé tel que succinctement exposé ci-  
20 dessus comporte une étape de transmission, à un système informatique distant relié audit  
réseau informatique, d'un ensemble d'information comportant une commande, un identifiant  
de l'utilisateur et un identifiant du boîtier et une étape de vérification, par le système  
informatique distant des droits associés auxdits identifiants et une étape de commande du  
boîtier, par le système informatique distant pour désactiver le filtrage et autoriser l'accès à  
25 tout contenu accessible sur le réseau informatique.

Grâce à ces dispositions, le fonctionnement du boîtier est plus sûr que si la décision  
de désactivation était effectuée uniquement par le boîtier, qui pourrait alors être piraté  
localement.

Selon des caractéristiques particulières, le procédé tel que succinctement exposé ci-  
30 dessus comporte, lorsque l'équipement a été désactivé, une étape d'activation de  
l'équipement lors du prochain redémarrage de l'ordinateur ou de la prochaine ouverture  
d'une session avec ledit ordinateur.

Selon un deuxième aspect, la présente invention vise un équipement, boîtier externe  
ou une carte interne à l'ordinateur, de filtrage de contenus en ligne, qui s'intercale entre  
35 l'ordinateur et un réseau informatique donnant accès à des contenus en ligne, ledit  
équipement recevant les contenus provenant du réseau, caractérisé en ce qu'il comporte :  
- un moyen d'analyse dudit contenu ;

- un moyen de recherche d'environnement dudit contenu sur ledit réseau ;
- un moyen d'analyse dudit environnement ;
- un moyen de décision de filtrage en fonction d'un ensemble de règles de décision dépendante des résultats des étapes d'analyse dudit contenu et de son environnement et
- 5 - un moyen de transmission ou non dudit contenu audit ordinateur, en fonction du résultat de l'étape de décision de filtrage.

Les avantages, buts et caractéristiques particulières de ce deuxième aspect étant identiques à ceux du procédé tel que succinctement exposé ci-dessus, ils ne sont pas rappelés ici.

10 D'autres avantages, buts et caractéristiques de la présente invention ressortiront de la description qui va suivre, faite dans un but explicatif et nullement limitatif en regard des dessins annexés dans lesquels :

- la figure 1 représente, schématiquement, le positionnement d'un boîtier conforme à la présente invention, dans un système informatique relié à un réseau informatique ;
- 15 - la figure 2 représente, schématiquement, les modules fonctionnels d'un mode de réalisation particulier du boîtier illustré en figure 1 et
- la figure 3 représente, schématiquement, un logigramme d'étapes mises en oeuvre dans un mode de réalisation particulier du procédé objet de la présente invention.

On observe, en figure 1, un ordinateur personnel (PC) 100, relié à un boîtier 110, lui-même relié à un modulateur-démodulateur (modem) 120 relié à un réseau informatique 130 lui-même relié à des serveurs distants 140, 150 et 160. Les liaisons représentées peuvent être filaires ou sans fil, selon des techniques de communication connues.

L'ordinateur personnel (PC) 100 représente un système informatique pouvant comporter un ordinateur personnel de type connu ou un réseau local de plusieurs ordinateurs de type connu. Au cours de l'installation de l'application informatique qui, dans l'ordinateur personnel 100 gère la communication avec le boîtier 110, on installe un pilote (en anglais driver) du boîtier pour que l'ordinateur personnel ne puisse pas accéder au réseau informatique 130 sans passer par l'intermédiaire du boîtier 110. Le fonctionnement du boîtier n'est donc pas désactivable comme un logiciel, il est intégré au fonctionnement de l'ordinateur 100 par une liaison sécurisée contrôlée en permanence.

30 Le boîtier 110, objet de la présente invention comporte une carte électronique 111 à microprocesseur 112 et à mémoire non volatile 113 et des interfaces 114 et 115 lui permettant de communiquer, d'une part, avec l'ordinateur personnel (PC) 100 et, d'autre part, avec le modem 120 et, par l'intermédiaire de ce modem 120 et le réseau informatique 130, avec les serveurs 140, 150 et 160.

La mémoire non volatile 113 conserve des instructions de programme destiné à être exécuté par le microprocesseur 112 pour mettre en oeuvre le procédé objet de la présente

invention, et, par exemple, les fonctions illustrées en figure 2 et/ou le logigramme illustré en figure 3.

Dans le mode de réalisation décrit en figure 1, le boîtier 110 comporte un moyen d'identification à clé matérielle 116, par exemple à carte à puce ou à mesure biométrique, par exemple un lecteur d'empreinte digitale.

Le modem 120 est de type connu, par exemple pour communication sur un réseau commuté, éventuellement en liaison haut débit. Le réseau informatique 130 est, par exemple, le réseau Internet. Les serveurs distants 140, 150 et 160 sont de type connu. Dans le mode de réalisation représenté ici, le serveur 140 est dédié au contrôle, au renseignement et à la commande de boîtiers identiques au boîtier 110. Dans d'autres modes de réalisation, le boîtier 110 ne fonctionne pas sous le contrôle d'un serveur distant.

Le serveur 140 conserve tout ou partie des bases de données mises en oeuvre par les boîtiers 110, par exemple des dictionnaires de mots et chaque boîtier 110 met à jour ses bases de données, par référence aux bases de données conservées par le serveur 140.

Les serveurs 150 et 160 conservent un contenu d'information. Par exemple, le serveur 150 est un serveur hébergeant un site commercial de vente d'appareils électroménagers, un site d'information sur les brevets et un site médical traitant de différentes pathologies du corps humain et le serveur 160 est un serveur hébergeant un site pour adultes comportant des contenus, en particulier des images et des films comportant des images, à caractère pornographique.

En variante, le boîtier 110 est remplacé par une carte interne à l'ordinateur personnel 100 et fonctionne comme décrit ci-dessus. Dans la suite de la description, le terme "boîtier" couvre aussi bien le cas d'un boîtier externe à l'ordinateur personnel 100 que le cas d'une carte électronique interne à l'ordinateur personnel 100.

On observe que le boîtier 110 peut, en variante, être placé entre le modem 120 et le réseau informatique 130. Dans ce cas, il comporte, lui-même un modem pour communiquer sur le réseau informatique 130.

Le boîtier 110 contient différents modules qui interagissent ensemble pour créer un système efficace de filtrage de données entrant dans l'ordinateur et, éventuellement, un coupe-feu (en anglais firewall), un module anti-virus, un module anti-fenêtre surgissante (en anglais popup), un module anti-pourriel (en anglais spam), ces modules utilisant les ressources de calcul et de mémoire du boîtier 110 sans consommer les ressources de l'ordinateur personnel 100 et évitent ainsi que les virus atteignent l'ordinateur personnel 100.

Pour installer le boîtier 110, dans l'une des configurations illustrées en figure 1, on procède de la manière suivante :

- connecter le boîtier entre le modem et l'ordinateur ;

- identifier ou authentifier, par le moyen d'identification à clé matérielle 116 du boîtier 110, la personne qui aura les droits de désactiver ou de désinstaller le boîtier, soit par insertion d'une clé matérielle, soit par reconnaissance d'une mesure biométrique, par exemple par le lecteur d'empreinte digitale ;

5 - effectuer l'installation, par exemple, en accédant au serveur 140, ou en insérant un disque compact (CD-ROM) dans le lecteur de CD-ROM de l'ordinateur 100 et démarrer l'installation, au cours de l'installation, l'utilisateur autorisé indique s'il souhaite recevoir un courrier électronique à chaque fois qu'il y aura désactivation du boîtier 110 et, si oui, à quelle adresse électronique il souhaite recevoir les courriers électroniques correspondants ;

10 - le boîtier 110 identifie alors l'ordinateur 100, c'est-à-dire en détermine un profil suffisamment unique pour reconnaître l'ordinateur 100 lors d'utilisations postérieures, se connecte au serveur distant 140 et lui fournit un identifiant (par exemple un numéro de série qu'il conserve en mémoire non volatile) ;

- le serveur distant 140 vérifie alors le bon fonctionnement du boîtier 110, vérifie la validité de l'abonnement de l'utilisateur dudit boîtier et initialise le boîtier. L'utilisateur enregistre alors son code d'identification personnel ou enregistre l'empreinte digitale de l'utilisateur désigné, c'est-à-dire d'un adulte, qui authentifie l'utilisateur désigné (sert aussi d'identification pour l'accès aux données en ligne concernant le fonctionnement du boîtier et l'abonnement aux services de protection qu'il effectue) ;

15 - une étape supplémentaire est ajoutée à la procédure de démarrage de l'ordinateur 100 : la vérification de la présence du boîtier 110 sans lequel l'accès à Internet n'est pas autorisé, donc impossible ;

- le filtrage est alors activé par défaut à chaque redémarrage de l'ordinateur 100 ou à chaque ouverture d'une session avec l'ordinateur 100, la désactivation du boîtier 110 ou le changement de ses paramètres nécessitant l'identification de la personne autorisée par le moyen d'identification à clé matérielle 116.

20 Pour la suite du fonctionnement, l'ordinateur personnel 100 et le boîtier 110 effectuent une vérification de la présence du boîtier 110 et de l'ordinateur personnel 100, respectivement, et en cas de détection d'absence, ils envoient un signal de détection d'absence au serveur distant 140 et courrier électronique à l'utilisateur identifié par le boîtier 110 puis coupent la connexion au réseau informatique 130 et bloquent la possibilité de connexion au réseau informatique 130.

Après authentification de l'identité de l'utilisateur, il est possible de désactiver, désinstaller ou modifier les paramètres de filtrage du boîtier 110 :

35 - interdire les téléchargements de certains types de fichier (".mpeg", ".avi", ".zip" ...),

- bloquer les sites peer-to-peer (en français "poste à poste"),

- bloquer les chats synchrones ou, au moins le transfert de documents sur ces chats sauf si le chat met en oeuvre des identifications par adresse de courrier électronique et si l'adresse du correspondant correspond à une adresse présente dans un carnet d'adresses électroniques référencée comme "fiable" par l'utilisateur autorisé du boîtier 110,

- 5 - bloquer NNTP (newsgroup ou groupe de discussion) et/ou  
- ne pas analyser les courriers électroniques entrants venant d'adresses considérées comme fiables dans le carnet d'adresses lié aux fonctions de filtrage.

Chaque désactivation du boîtier donne lieu à la transmission, au serveur 140, d'une information témoin afin que le serveur 140 conserve une trace de cette désactivation que  
10 l'utilisateur pourra consulter après avoir été identifié par le moyen d'identification à clé matérielle 116.

On observe, en figure 2, une entrée 200 d'information en provenance du réseau 130, un module d'acquisition et de tri de type d'information 210, un module de traitement contextuel 220, un module de traitement sémantique et textuel 230, un module de décision  
15 240 comportant un module de première décision 241 et un module de deuxième décision 242, un module d'analyse d'images 250, une sortie d'information 260 à destination de l'ordinateur 100 et un module 270 de transmission d'information sur le réseau 130.

L'entrée 200 reçoit toute l'information en provenance du réseau 130 destinée à l'ordinateur 100, sous forme de trame conforme au protocole IP (pour Internet Protocol). Le  
20 module d'acquisition et de tri de type d'information 210 reçoit cette information et la trie selon son type :

- information en provenance d'un site web,
  - information en provenance d'un site de chat,
  - information de courrier électronique,
- 25 en fonction du protocole selon lequel cette information est transmise (respectivement les protocoles HTTP, NNTP et SMTP ou autre).

D'une manière générale et préférentiellement, le boîtier 110 effectue le filtrage des données en effectuant d'abord des analyses qui peuvent être très rapides (analyse de mots clés et de balises, par exemple) et s'il peut conclure de cette première analyse que  
30 l'information ne doit pas être transmise à l'utilisateur du PC, il ne la transmet pas et, dans le cas contraire, il effectue une deuxième analyse qui prend plus de temps de traitement (traitement des pages liées à la page analysée, des critères de la page, voir plus loin, des javascripts, ...) et s'il peut conclure de cette deuxième analyse que l'information ne doit pas être transmise à l'utilisateur du PC, il ne la transmet pas et, dans le cas contraire, il effectue  
35 une troisième analyse (par exemple traitement d'images de la page exposé ci-dessous) et ainsi de suite jusqu'à ce que tous les traitements aient été effectués et que la dernière décision de transmission ou non, de la page, ait été prise.

Dans un but de simplification, on ne décrit, ci-dessous, que deux étapes et moyens de traitement suivies de deux étapes et moyens de décision.

Le module de traitement contextuel 220 détermine et traite les informations suivantes :

- 5 a) S'il s'agit d'une information en provenance d'un site web (protocole HTTP) le module de traitement contextuel 220 analyse le contenu de la page reçue ;
  - il détermine la langue de la page, compare les mots clés contenus dans l'adresse électronique (URL) de la page, dans les métatags "keyword" et "description" et dans le code source de la page avec un dictionnaire de mots interdits les plus courants (dictionnaire
  - 10 conservé dans la mémoire non volatile du boîtier 110) ;
  - il recherche des balises spécifiques d'auto déclaration de contenu de la page (par exemple balises PICS, ICRA ...) ;
  - si la page demandée possède une adresse électronique (URL) qui ne correspond pas à la page d'accueil du site, il recherche cette page d'accueil sur le réseau 130 (en raccourcissant
  - 15 l'adresse électronique URL par retrait de ses derniers caractères, éventuellement en plusieurs étapes, et en fonction des caractères "/" et, dans cette page d'accueil, un avertissement ("disclaimer") en cas de caractère sensible de la page pouvant heurter qui demande une acceptation volontaire (par clic sur un bouton "entrer") ;
  - il effectue un relevé des différents critères de la page : nombre de mots, de liens
  - 20 hypertextes, d'images, scripts, tailles de fichiers, format des fichiers, scripts, contenu du texte et vecteurs sémantiques (groupement de mots ayant une signification)...
  - il analyse les javascripts (leur présence et leur action, par exemple ouverture de page ou popup et analyse de popup) ;
  - il recherche, télécharge et analyse les pages accessibles par les liens présents sur la page
  - 25 analysée comme indiqué ci-dessus ;

Dans un mode de réalisation préférentiel, le module de traitement contextuel 220 effectue une collecte des textes de la page au cours de laquelle, si des textes sont incrustés dans des graphismes ou images, ces textes en sont extraits et ajoutés aux informations de la page reçues au format texte, aux textes de l'adresse électronique (URL) de la page et des

30 métatags "keyword" et "description". Par exemple une reconnaissance optique de caractère est effectuée pour extraire les textes des images et graphismes.

b) s'il s'agit d'information de type courrier électronique (protocole SMTP), la philosophie de filtrage du courrier électronique est basée sur un confort de l'utilisateur qui ne sera pas agressé par des courriers non désirés (publicités, spam, mailings-listes automatiques,

35 contenu des pièces jointes). Si le courrier électronique entrant provient d'une adresse électronique (adresse e-mail) présente fiable dans le carnet d'adresses lié aux fonctions de filtrage, dans la mémoire du boîtier, le courrier n'est pas analysé. Si le courrier électronique



entrant ne provient pas d'un émetteur enregistré dans le carnet d'adresse, le module de traitement contextuel 220 :

- détermine s'il y a au moins une image ou un fichier susceptible d'en contenir une dans le corps du courrier électronique ou dans les fichiers attachés ;
- 5 - lit et analyse les liens contenus dans les courriers électroniques (et analyse des métatags de la page liée) comme indiqué ci-dessus ;
- effectue une analyse textuelle du contenu du mail comme indiqué ci-dessus.

Dans un mode de réalisation préférentiel, le module de traitement contextuel 220 effectue une simplification linguistique multilingue au cours de laquelle la langue des informations textuelles est d'abord déterminée, de manière connue, puis chaque mot de  
10 texte est mis en relation avec un synonyme dans la même langue, synonyme qui peut être le mot d'origine lui-même ou avec un mot de la même langue considéré comme ayant un sens voisin, en mettant en oeuvre une table de correspondance ou un dictionnaire de synonymes ou de mots possédant un sens voisin.

15 c) pour l'information provenant de chat ou de news groups (protocole NNTP), le module de traitement contextuel 220 détermine si les informations provenant des tiers proviennent d'utilisateurs référencées, par l'utilisateur autorisé du boîtier 110, comme fiables, dans le carnet d'adresses électroniques.

20 Les résultats des traitements effectués par le module de traitement contextuel 220 sont simultanément transmis au module de traitement sémantique et textuel 230 et au module de première décision 241.

Dans un mode de réalisation préférentiel, le module de traitement sémantique et textuel 230 détermine le type de contenu sémantique de la page par une analyse morpho-syntaxique du texte, en utilisant des vecteurs conceptuels (thésaurus et/ou dictionnaire). Les  
25 résultats des traitements effectués par le module de traitement sémantique et textuel 230 sont transmis au module de première décision 241.

Puis, le module de traitement 230 effectue une extraction de critères par vectorisation de la page, et classification d'après des classifieurs spécialisés par catégories ou domaines. A cet effet, le module de traitement 230 compte des éléments, images, mots prédéfinis après  
30 leur simplification linguistique, par exemple.

Le module de première décision 241 effectue une première détermination de décision de transmettre ou non le contenu de la page à l'ordinateur 100, en fonction des résultats provenant d'au moins le module 220 et éventuellement du module 230. Lorsque l'un des traitements effectués par l'un de ces modules 220 et 230 fournit, par un traitement par règles  
35 logiques (règles "expertes"), un résultat immédiatement interprétable pour bloquer la transmission du contenu, par exemple présence d'un avertissement, la première décision est

de bloquer le contenu. A défaut, la première décision de filtrage est prise par un réseau de neurones ou en logique floue, selon des techniques connues.

Dans un mode de réalisation préférentiel, dans le module de traitement sémantique et textuel 230, un classifieur secondaire traite les résultats pour chaque critère de tri (nombre  
5 d'images, nombre de mots prédéfinis, par exemple) et fournit un résultat de classification ou note et un classifieur traite les résultats des classifieurs secondaires, en les pondérant éventuellement, pour déterminer si la page peut être transmise à l'utilisateur.

Le résultat de la première décision peut être :

- décision de blocage du contenu,
- 10 - décision de faire parvenir le contenu à l'ordinateur 100,
- décision de poursuivre l'analyse de ce contenu.

Dans le troisième cas, l'information à traiter est transmise au module d'analyse d'image 250, qui effectue les traitements suivants :

- extraction de caractères et reconnaissance des mots dans les fichiers images (par exemple  
15 boutons, images et graphismes) présents sur la page, par exemple avec reconnaissance optique de caractères ;
- transmission de ces mots au module de traitement contextuel 220 et au module de traitement sémantique 230 pour que les traitements exposés ci-dessus soit effectués ;
- recherche de texture chair (identifiée par la présence de peu de contours dans une couleur  
20 correspondant à de la chair et par une faible, mais non nulle, densité de points de contour sur la partie de couleur chair) dans les images, détermination du nombre d'images en contenant ;
- relevé de contours des zones contenant la texture chair, reconnaissance de formes, recherche des yeux, de la bouche, des mains dans l'image pour déterminer les postures des  
25 différents sujets, nombre de sujets dans l'image, gros plans (ces étapes peuvent être effectuées par un réseau neuronal) ;
- dans le cas des courriers électroniques, des newsgroups et des chats, analyse des fichiers images joints et
- il analyse d'autres éléments de l'environnement de la page (bannières, images  
30 surgissantes ou "pop-up"), comme indiqué ci-dessus.

En fonction des résultats de ces traitement, le module de deuxième décision 242 prend une décision finale, par la mise en oeuvre d'un réseau de neurones ou de logique floue :

- décision de blocage du contenu, en fonction des paramètres personnalisés par  
35 l'utilisateur ou
- décision de faire parvenir le contenu à l'ordinateur 100.

On observe que le module de deuxième décision 242 peut, par exemple, mettre en oeuvre un classifieur de Bayes et un arbre de décision (cette méthode étant réputée fiable éprouvée et rapide).

5 En variante, le module de deuxième décision effectue les mêmes traitements que le module de première décision mais appliqué à l'environnement de la page, par exemple d'autres pages vers lesquelles mènent les liens présents sur la page et la décision finale de transmission à l'utilisateur est prise sur lesquels les modules 220 et 230 sont mis en oeuvre.

La sortie d'information 260 à destination de l'ordinateur 100 permet, lorsque l'image n'est pas filtrée ou bloquée, de transmettre le contenu de la page demandée à l'ordinateur  
10 100.

Lorsque l'utilisateur désigné souhaite arrêter le fonctionnement du boîtier 110, le module 270 de transmission d'information sur le réseau transmet au serveur distant 140, un triplet d'information comportant la commande de l'utilisateur, son identifiant et celui du boîtier 110. Le serveur distant 140 vérifie les droits et les informations transmises et commande  
15 éventuellement le boîtier 110, pour autoriser l'accès à tout contenu accessible sur le réseau 130.

On rappelle, ci-dessous, l'approche floue de l'analyse ou de la classification.

Les modèles flous ou systèmes d'inférence floue (SIF) permettent de représenter le comportement de systèmes complexes. La théorie des ensembles flous permet une  
20 représentation simple des incertitudes et imprécisions liées aux informations et aux connaissances. Son principal avantage est d'introduire le concept d'appartenance graduelle à un ensemble alors qu'en logique ensembliste classique, cette appartenance est binaire (appartient ou n'appartient pas à un ensemble). Un élément peut ainsi appartenir à plusieurs ensembles avec des degrés d'appartenance respectifs de 0,15 et 0,6, par exemple.

25 On observe, en figure 3, une succession d'étapes mises en oeuvre dans un mode particulier de réalisation du procédé objet de la présente invention.

A la suite de l'étape 300 d'initialisation de l'ordinateur 100 et du boîtier 110, au cours d'une étape 302, l'ordinateur 100 détermine si le boîtier 110 lui est correctement relié. Si non, l'ordinateur 100 interdit toute connexion au réseau informatique 130 et le processus de  
30 fonctionnement selon le procédé objet de la présente invention est achevé. Ainsi, à chaque redémarrage de l'ordinateur et à chaque ouverture d'une session sur cet ordinateur, on active l'équipement pour filtrer les contenus accessibles en ligne.

Si le boîtier 110 est correctement relié à l'ordinateur, au cours d'une étape 304, on détermine si l'utilisateur tente d'accéder à un contenu en ligne. Si non, on retourne à l'étape  
35 304. Si oui, au cours d'une étape 306, le boîtier autorise la connexion au réseau 140 et détermine si l'utilisateur a entré une commande de désactivation. Si non, on passe à l'étape 314. Si oui, au cours d'une étape 308, on vérifie l'identité de l'utilisateur désigné, par

exemple en identifiant une clé matérielle (par exemple une carte à mémoire ou une empreinte digital) et on transmet au serveur distant 140, un triplet d'information comportant la commande de l'utilisateur, son identifiant et celui du boîtier 110. Le serveur distant 140 vérifie les droits et les informations transmises, étape 310, et si l'utilisateur désigné est  
5 authentifié, il commande le boîtier 110, pour autoriser l'accès à tout contenu accessible sur le réseau 130, étape 312, et le processus de fonctionnement selon le procédé objet de la présente invention est achevé.

Au cours de l'étape 314, on trie l'information en provenance du réseau informatique 130 selon son type :

- 10 - information en provenance d'un site web,
- information en provenance d'un site de chat,
- information de courrier électronique,

en fonction du protocole selon lequel cette information est transmise (respectivement les protocoles HTTP, NNTP et SMTP).

15 Au cours d'une étape 316, on détermine et traite les informations suivantes :

a) S'il s'agit d'une information en provenance d'un site web (protocole HTTP), on analyse le contenu de la page reçue ;

- on détermine la langue de la page, compare les mots clés contenus dans l'adresse électronique (URL) de la page, dans les métatags "keyword" et "description" et dans le code  
20 source de la page avec un dictionnaire de mots interdits les plus courants (dictionnaire conservé dans la mémoire non volatile du boîtier 110) ;

- on recherche des balises spécifiques d'auto déclaration de contenu de la page (par exemple balises PICS, ICRA ...) ;

- si la page demandée possède une adresse électronique (URL) qui ne correspond pas à la  
25 page d'accueil du site, on recherche cette page d'accueil sur le réseau 130 (en raccourcissant l'adresse électronique URL par retrait de ses derniers caractères, éventuellement en plusieurs étapes, et en fonction des caractères "/" ) et, dans cette page d'accueil, un avertissement ("disclaimer") en cas de caractère sensible de la page pouvant heurter qui demande une acceptation volontaire (par clic sur un bouton "entrer") ;

30 - on effectue un relevé des différents critères de la page : nombre de mots, de liens hypertextes, d'images, scripts, tailles de fichiers, format des fichiers, scripts, contenu du texte et vecteurs sémantiques (groupement de mots ayant une signification)...

- on analyse les javascripts (leur présence et leur action, par exemple ouverture de page ou popup et analyse de popup) ;

35 - on recherche, télécharge et analyse les pages accessibles par les liens présents sur la page analysée comme indiqué ci-dessus ;

b) s'il s'agit d'information de type courrier électronique (protocole SMTP), la philosophie de filtrage du courrier électronique est basée sur un confort de l'utilisateur qui ne sera pas agressé par des courriers non désirés (publicités, spam, mailings-listes automatiques, contenu des pièces jointes). Si le courrier électronique entrant provient d'une adresse électronique (adresse e-mail) présente fiable dans le carnet d'adresses lié aux fonctions de filtrage, dans la mémoire du boîtier, le courrier n'est pas analysé. Si le courrier électronique entrant ne provient pas d'un émetteur enregistré dans le carnet d'adresse, on :

- détermine s'il y a au moins une image ou un fichier susceptible d'en contenir une dans le corps du courrier électronique ou dans les fichiers attachés ;

- lit et analyse les liens contenus dans les courriers électroniques (et analyse des métatags de la page liée) comme indiqué ci-dessus ;

- effectue une analyse textuelle du contenu du mail comme indiqué ci-dessus.

Dans un mode de réalisation préférentiel, au cours de l'étape 316, on effectue une collecte des textes de la page au cours de laquelle, si des textes sont incrustés dans des graphismes ou images, ces textes en sont extraits et ajoutés aux informations de la page reçues au format texte. Par exemple une reconnaissance optique de caractère est effectuée pour extraire les textes des images et graphismes.

En cas de filtrage, on avertit l'utilisateur de l'ordinateur personnel, par ouverture d'une boîte de dialogue, et on ne détruit pas les fichiers.

c) pour l'information provenant de chat ou de news groups (protocole NNTP), on détermine si les informations provenant des tiers proviennent d'utilisateurs référencées, par l'utilisateur autorisé du boîtier 110, comme fiables, dans le carnet d'adresses électroniques.

Puis, au cours d'une étape 318, on détermine le type de contenu sémantique de la page par une analyse morpho-syntaxique du texte, en utilisant des vecteurs conceptuels (thésaurus et/ou dictionnaire).

Dans un mode de réalisation préférentiel, au cours de l'étape 318, on effectue une simplification linguistique multilingue au cours de laquelle la langue des informations textuelles est d'abord déterminée, de manière connue, puis chaque mot de texte est mis en relation avec un synonyme dans la même langue, synonyme qui peut être le mot d'origine lui-même ou avec un mot de la même langue considéré comme ayant un sens voisin, en mettant en oeuvre une table de correspondance ou un dictionnaire de synonymes ou de mots possédant un sens voisin.

Dans ce mode de réalisation préférentiel, au cours de l'étape 318, on effectue une extraction de critères par vectorisation de la page, et classification d'après des classifieurs spécialisés par catégories ou domaines. A cet effet, le module de traitement 230 compte des éléments, images, mots prédéfinis après leur simplification linguistique, par exemple.

Au cours d'une étape 320 de détermination de première décision, on effectue une première détermination de décision de transmettre ou non le contenu de la page à l'ordinateur 100, en fonction des résultats provenant des étapes 316 et 318.

Lorsque l'un des traitements effectués par l'un de ces modules fournit, par un traitement par règles logiques, un résultat immédiatement interprétable pour bloquer la transmission du contenu, par exemple présence d'un avertissement, au cours de l'étape 320, on détermine que la première décision est de bloquer le contenu. Dans un mode de réalisation préférentiel, au cours de l'étape 320, un classifieur secondaire traite les résultats pour chaque critère de tri (nombre d'images, nombre de mots prédéfinis, par exemple) et fournit un résultat de classification ou note et un classifieur traite les résultats des classifieurs secondaires, en les pondérant éventuellement, pour déterminer si la page peut être transmise à l'utilisateur.

A défaut, la première décision de filtrage est prise par un réseau de neurones ou en logique floue, selon des techniques connues. Le résultat de cette première décision peut être :

- décision de blocage du contenu (on ne fournit pas le contenu à l'ordinateur et on affiche un message de refus d'accès, étape 322) ;
- décision de faire parvenir le contenu à l'ordinateur 100 (on fournit le contenu à l'ordinateur 100 comme si le boîtier 110 n'était pas associé à l'ordinateur, étape 324) ou
- décision de poursuivre l'analyse de ce contenu.

Dans le troisième cas, au cours d'une étape 326, on effectue les traitements suivants :

- extraction de caractères et reconnaissance des mots dans les fichiers images (par exemple boutons, images et graphismes) présents sur la page, par exemple avec reconnaissance optique de caractères ;
- traitement contextuel comme indiqué à l'étape 316 et traitement sémantique comme indiqué à l'étape 318 ;
- recherche de texture chair (identifiée par la présence de peu de contours dans une couleur correspondant à de la chair et par une faible, mais non nulle, densité de points de contour sur la partie de couleur chair) dans les images, détermination du nombre d'images en contenant ;
- relevé de contours des zones contenant la texture chair, reconnaissance de formes, recherche des yeux, de la bouche, des mains dans l'image pour déterminer les postures des différents sujets, nombre de sujets dans l'image, gros plans (ces étapes peuvent être effectuées par un réseau neuronal) ;
- dans le cas des courriers électroniques, des newsgroups et des chats, analyse des fichiers images joints et

- analyse d'autres éléments de l'environnement de la page (bannières, fenêtres surgissantes ou "pop-up"), comme indiqué ci-dessus.

En fonction des résultats de ces traitements, au cours d'une étape de deuxième décision 328, on prend une décision finale, par la mise en oeuvre d'un réseau de neurones ou de logique floue :

- décision de blocage du contenu, étape 322, en fonction des paramètres personnalisés par l'utilisateur ou

- décision de faire parvenir le contenu à l'ordinateur 100, étape 324.

A la suite de l'une des étapes 322 ou 324, on retourne à l'étape 314.

En variante, l'étape 328 effectue les mêmes traitements que ceux appliqués pour la première décision mais appliqués à l'environnement de la page, par exemple d'autres pages vers lesquelles mènent les liens présents sur la page et la décision finale de transmission à l'utilisateur est prise sur lesquels les modules 220 et 230 sont mis en oeuvre.

En variante, l'étape de validation de commande de l'utilisateur est effectuée dès que l'utilisateur est authentifié, par mot de passe ou mesure biométrique, par exemple, sans recours au serveur distant 140.

En variante, l'étape 318 est supprimée.

On observe que la deuxième décision, étape 328, peut, par exemple, mettre en oeuvre un classifieur de Bayes et un arbre de décision (cette méthode étant réputée fiable éprouvée et rapide).

Préférentiellement, la classification se fait après un apprentissage "en laboratoire" de catégories de pages, selon des techniques connues dans le domaine de la recherche de contenu (en anglais "web mining" ou "content mining"). A cet effet, on donne au classifieur des grandes quantités de pages de chaque catégorie à apprendre et il reconnaît ensuite automatiquement à quelle catégorie appartient une nouvelle page qui lui est soumise.

**REVENDEICATIONS :**

- 1 - Procédé de filtrage de contenus en ligne, caractérisé en ce qu'il comporte :
- de mettre en oeuvre un équipement (110), boîtier externe ou une carte interne à l'ordinateur, qui s'intercale entre l'ordinateur (100) et un réseau informatique (130) donnant accès à des contenus en ligne, ledit équipement recevant le contenu provenant du réseau ;
  - une étape d'analyse (316, 318, 326) dudit contenu ;
  - une étape de recherche d'environnement (316, 318, 326) dudit contenu sur ledit réseau ;
  - une étape d'analyse (316, 318, 326) dudit environnement ;
  - une étape de décision de filtrage (320, 328) en fonction d'un ensemble de règles de décision dépendante des résultats des étapes d'analyse dudit contenu et de son environnement et
  - une étape de transmission (324) ou non dudit contenu audit ordinateur, en fonction du résultat de l'étape de décision de filtrage.
- 2 - Procédé selon la revendication 1, caractérisé en ce que, au cours de l'étape d'analyse dudit environnement (316, 318, 326), on traite les pages vers lesquelles mènent les liens hypertextes dudit contenu.
- 3 - Procédé selon l'une quelconque des revendications 1 ou 2, caractérisé en ce qu'au moins une l'étape d'analyse (316, 318, 326) dudit contenu comporte une première étape de tri rapide de contenu (316, 318), l'étape de décision comportant une première étape de détermination de décision (320) en fonction du résultat de ladite première étape de tri rapide et, en cas d'indétermination du résultat de ladite première étape de détermination de décision, l'étape d'analyse comporte une deuxième étape de tri de contenu (326) d'une durée supérieure à la première étape de tri rapide, l'étape de décision comportant alors une deuxième étape de détermination de décision (328) en fonction du résultat de la deuxième étape de tri.
- 4 - Procédé selon la revendication 3, caractérisé en ce que la première étape de tri rapide de contenu (316, 318) traite un contenu ne comportant pas d'image et en ce que la deuxième étape de tri de contenu (326) comporte une étape de traitement d'image.
- 5 - Procédé selon l'une quelconque des revendications 1 à 4, caractérisé en ce qu'au moins une étape d'analyse (326) comporte une l'étape de traitement d'image au cours de laquelle, pour au moins une image, on analyse la texture du contenu de l'image pour extraire les parties de l'image dont la texture correspond à de la chair.



6 - Procédé selon la revendication 5, caractérisé en ce que l'étape de traitement d'image comporte une étape d'analyse de posture du ou des personnages dont on voit des parties charnelles.

5 7 - Procédé selon l'une quelconque des revendications 1 à 6, caractérisé en ce qu'au moins une étape d'analyse (326) comporte une étape d'extraction de caractères d'images incorporées dans le contenu en ligne.

10 8 - Procédé selon l'une quelconque des revendications 1 à 7, caractérisé en ce qu'il comporte une étape d'identification de l'utilisateur (308, 310) et une étape de désactivation (312) de filtrage et d'autorisation d'accès à tout contenu accessible sur le réseau informatique, en fonction du résultat de ladite identification.

15 9 - Procédé selon l'une quelconque des revendications 1 à 8, caractérisé en ce qu'il comporte une étape de transmission (310), à un système informatique distant (140) relié audit réseau informatique (130), d'un ensemble d'information comportant une commande, un identifiant de l'utilisateur et un identifiant de l'équipement et une étape de vérification, par le système informatique distant des droits associés auxdits identifiants et une étape de commande de l'équipement (310, 312), par le système informatique distant pour désactiver  
20 le filtrage et autoriser l'accès à tout contenu accessible sur le réseau informatique.

10 - Procédé selon l'une quelconque des revendications 8 ou 9, caractérisé en ce qu'il comporte, lorsque l'équipement (110) a été désactivé, une étape d'activation (300) de l'équipement lors du prochain redémarrage de l'ordinateur ou de la prochaine ouverture  
25 d'une session avec ledit ordinateur.

11 - Equipement (110), boîtier externe ou une carte interne à un ordinateur (100), de filtrage de contenus en ligne, qui s'intercale entre l'ordinateur et un réseau informatique (130) donnant accès à des contenus en ligne, ledit équipement recevant les contenus provenant  
30 du réseau, caractérisé en ce qu'il comporte :

- un moyen d'analyse (112, 113) dudit contenu ;
- un moyen de recherche (112, 113) d'environnement dudit contenu sur ledit réseau ;
- un moyen d'analyse (112, 113) dudit environnement ;
- un moyen de décision de filtrage (112, 113) en fonction d'un ensemble de règles de  
35 décision dépendante des résultats des étapes d'analyse dudit contenu et de son environnement et

- un moyen de transmission (114) ou non dudit contenu audit ordinateur, en fonction du résultat de l'étape de décision de filtrage.

5 12 - Equipement selon la revendication 11, caractérisé en ce que le moyen d'analyse (112, 113) dudit environnement traite les pages vers lesquelles mènent les liens hypertextes dudit contenu.

10 13 - Equipement selon l'une quelconque des revendications 11 ou 12, caractérisé en ce qu'au moins un moyen d'analyse dudit contenu (112, 113) est adapté à effectuer un premier tri rapide de contenu, le moyen de décision étant adapté à effectuer une première détermination de décision en fonction du résultat dudit premier tri rapide et, en cas d'indétermination du résultat de ladite première étape de détermination de décision, le moyen d'analyse est adapté à effectuer un deuxième tri de contenu d'une durée supérieure au premier tri rapide, le moyen de décision effectuant alors une deuxième détermination de  
15 décision en fonction du résultat du deuxième tri.

14 - Equipement selon la revendication 13, caractérisé en ce que le premier tri rapide de contenu traite un contenu ne comportant pas d'image et en ce que le deuxième tri de contenu comporte un traitement d'image.

20

15 - Equipement selon l'une quelconque des revendications 11 à 14, caractérisé en ce qu'au moins un moyen d'analyse (112, 113) comporte un moyen de traitement d'image adapté, pour au moins une image, à analyser la texture du contenu de l'image pour extraire les parties de l'image dont la texture correspond à de la chair.

25

16 - Equipement selon la revendication 15, caractérisé en ce que le traitement d'image comporte une analyse de posture du ou des personnages dont on voit des parties charnelles.

30 17 - Equipement selon l'une quelconque des revendications 11 à 16, caractérisé en ce qu'au moins un moyen d'analyse (112, 113) est adapté à extraire des caractères d'images incorporées dans le contenu en ligne.

35 18 - Equipement selon l'une quelconque des revendications 11 à 17, caractérisé en ce qu'il comporte un moyen d'identification de l'utilisateur à clé matérielle (116), le moyen de décision (112, 113) étant adapté, en fonction du résultat de l'identification, à désactiver le filtrage et à autoriser l'accès à tout contenu accessible sur le réseau informatique.

19 - Equipement selon l'une quelconque des revendications 11 à 18, caractérisé en ce qu'il comporte un moyen de transmission (115), à un système informatique distant (140) relié audit réseau informatique (130), d'un ensemble d'information comportant une commande, un identifiant de l'utilisateur et un identifiant de l'équipement et un moyen de réception, depuis le  
5 système informatique distant d'une commande de l'équipement pour désactiver le filtrage et autoriser l'accès à tout contenu accessible sur le réseau informatique.

20 - Equipement (110) selon l'une quelconque des revendications 18 ou 19, caractérisé en ce qu'il comporte un moyen d'activation (112, 113) adapté, lorsque l'équipement a été  
10 désactivé, à activer l'équipement lors du prochain redémarrage de l'ordinateur ou de la prochaine ouverture d'une session avec ledit ordinateur.

1/3

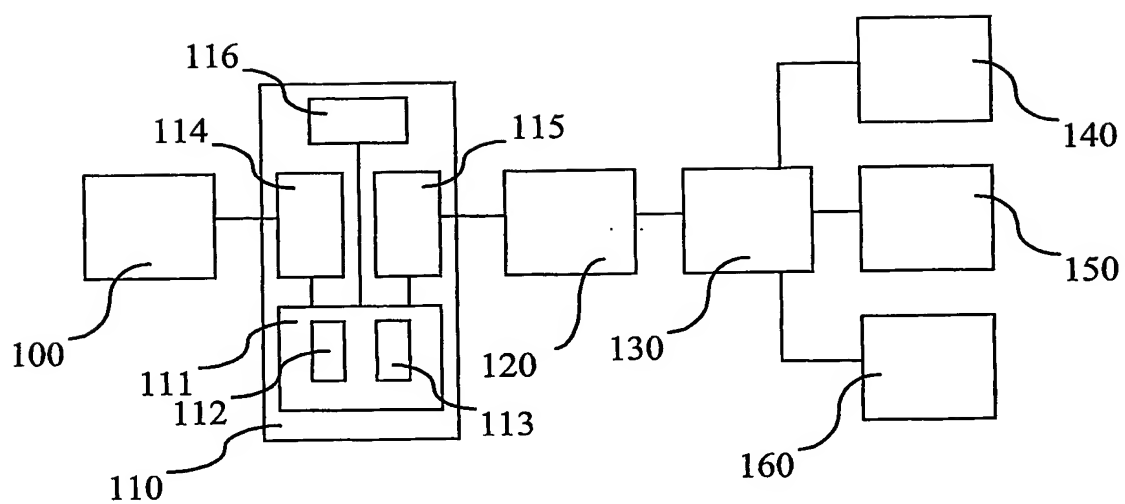


Figure 1

2/3

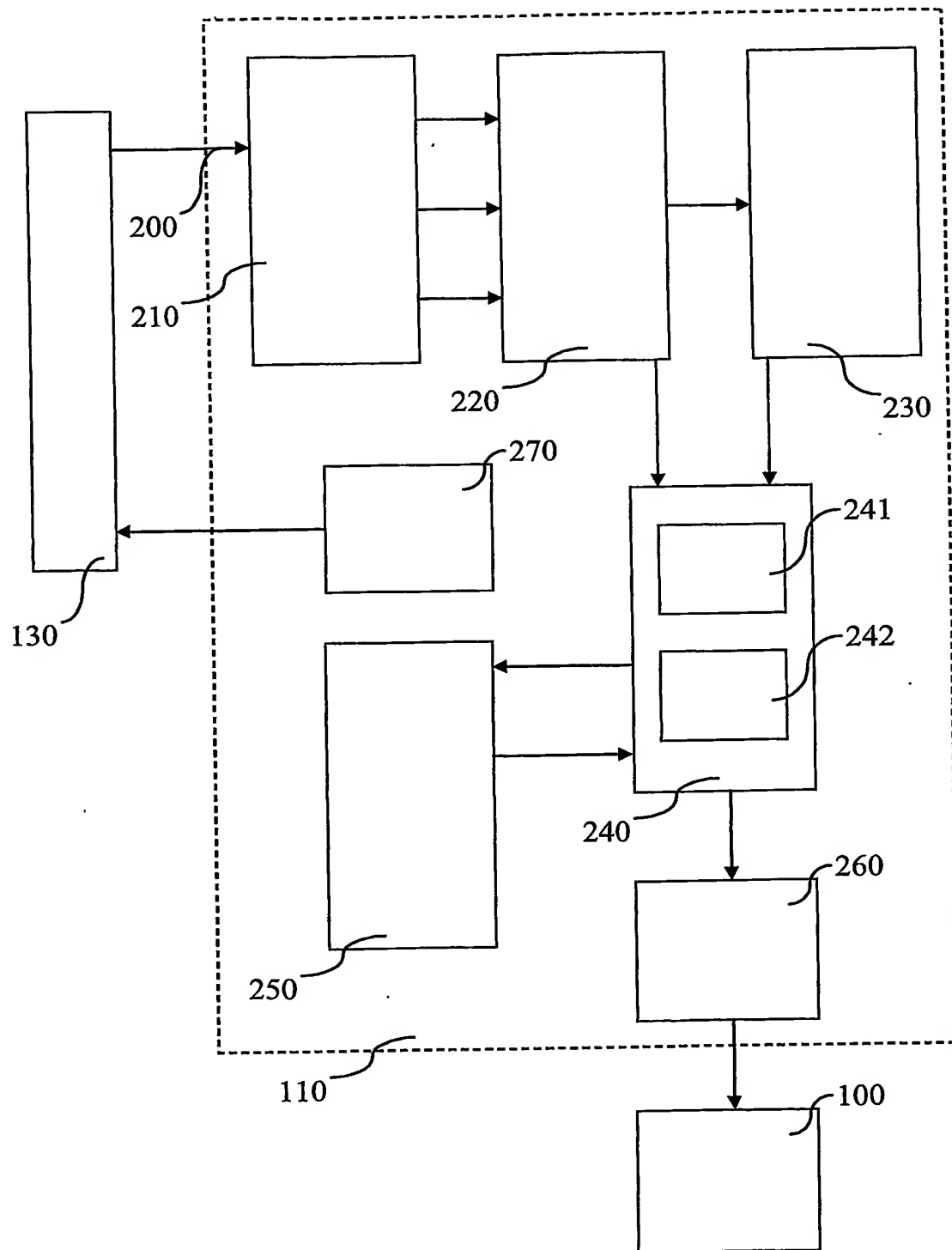


Figure 2

3/3

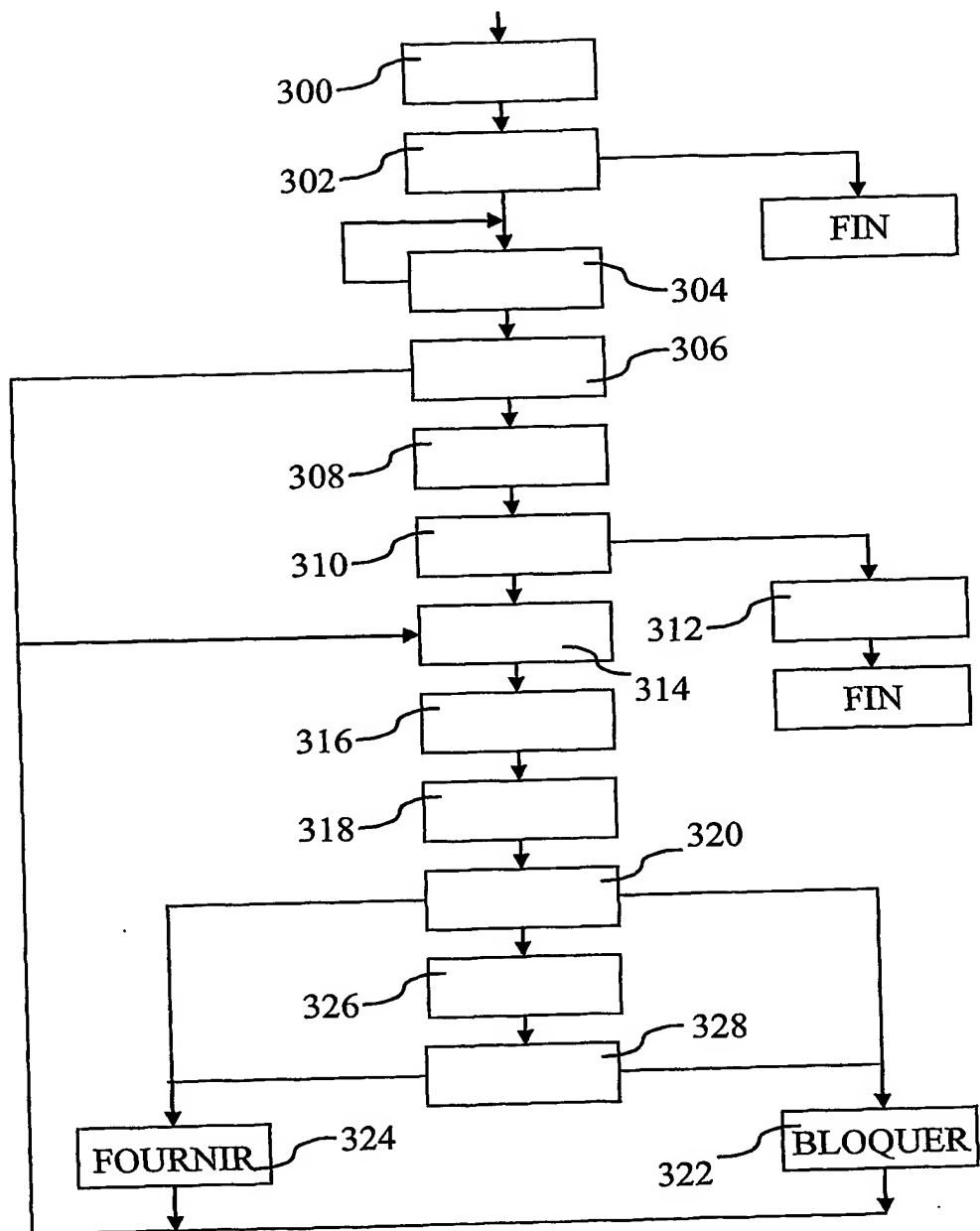


Figure 3

# INTERNATIONAL SEARCH REPORT

PCT/EP2004/052571

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 7 G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
IPC 7 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the International search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, IBM-TDB

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>GREENFIELD P ET AL: "Access Prevention techniques for Internet Content Filtering" INTERNET, 'Online! December 1999 (1999-12), XP002265027 Retrieved from the Internet: URL: <a href="http://www.cmis.csiro.au/Reports/filtering.pdf">http://www.cmis.csiro.au/Reports/filtering.pdf</a> 'retrieved on 2003-12-15! figures 2,5 page 9 page 19 - page 20</p> <p style="text-align: center;">----- -/--</p>	<p>1,3-7, 11,13-17</p>

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

### \* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*G\* document member of the same patent family

Date of the actual completion of the international search

9 February 2005

Date of mailing of the international search report

16/02/2005

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax (+31-70) 340-3016

Authorized officer

Michalski, S

# INTERNATIONAL SEARCH REPORT

PCT/EP2004/052571

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CHEN DING ET AL: "Centralized content-based Web filtering and blocking: how far can it go?" SYSTEMS, MAN, AND CYBERNETICS, 1999. IEEE SMC '99 CONFERENCE PROCEEDINGS. 1999 IEEE INTERNATIONAL CONFERENCE ON TOKYO, JAPAN 12-15 OCT. 1999, PISCATAWAY, NJ, USA, IEEE, US, 12 October 1999 (1999-10-12), pages 115-119, XP010363557 ISBN: 0-7803-5731-0 page 116, left-hand column page 117, left-hand column page 119, left-hand column	1-4, 11-14
X	US 2003/126267 A1 (BRODSKY TOMAS ET AL) 3 July 2003 (2003-07-03)  figures 1,4	1,3-5,7, 8,10,11, 13-15, 17,18,20
Y	paragraphs '0002!, '0007!, '0013!, '0015!, '0023!	2,12 5,6,15, 16
X	WO 01/98934 A (INFONUTZ LLC ;MORTL WILLIAM M (US)) 27 December 2001 (2001-12-27)  page 3 page 9 claim 1	1,3,4, 8-11,13, 14,18-20
Y	FORSYTH D A ET AL: "Identifying nude pictures" APPLICATIONS OF COMPUTER VISION, 1996. WACV '96., PROCEEDINGS 3RD IEEE WORKSHOP ON SARASOTA, FL, USA 2-4 DEC. 1996, LOS ALAMITOS, CA, USA, IEEE COMPUT. SOC, US, 2 December 1996 (1996-12-02), pages 103-108, XP010206418 ISBN: 0-8186-7620-5 abstract page 104, left-hand column, paragraphs 1,2	5,6,15, 16
A	"LOOK AHEAD FILTERING OF INTERNET CONTENT" IBM TECHNICAL DISCLOSURE BULLETIN, IBM CORP. NEW YORK, US, vol. 40, no. 12, 1 December 1997 (1997-12-01), page 143, XP000754118 ISSN: 0018-8689	1,11
Y	the whole document	2,12



# INTERNATIONAL SEARCH REPORT

PCT/EP2004/052571

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
US 2003126267	A1	03-07-2003	AU	2002367040 A1	30-07-2003
			WO	03060757 A2	24-07-2003
WO 0198934	A	27-12-2001	AU	6857901 A	02-01-2002
			WO	0198934 A2	27-12-2001

# RAPPORT DE RECHERCHE INTERNATIONALE

PCT/EP2004/052571

A. CLASSEMENT DE L'OBJET DE LA DEMANDE  
CIB 7 G06F17/30

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)  
CIB 7 G06F

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si réalisable, termes de recherche utilisés)  
EPO-Internal, WPI Data, PAJ, IBM-TDB

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie *	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	<p>GREENFIELD P ET AL: "Access Prevention techniques for Internet Content Filtering" INTERNET, 'Online! décembre 1999 (1999-12), XP002265027</p> <p>Extrait de l'Internet: URL: <a href="http://www.cmis.csiro.au/Reports/filtering.pdf">http://www.cmis.csiro.au/Reports/filtering.pdf</a> &gt; 'extrait le 2003-12-15! figures 2,5 page 9 page 19 - page 20</p>	<p>1,3-7, 11,13-17</p>

☒ Voir la suite du cadre C pour la fin de la liste des documents

☒ Les documents de familles de brevets sont indiqués en annexe

\* Catégories spéciales de documents cités:

- \*A\* document définissant l'état général de la technique, non considéré comme particulièrement pertinent
- \*E\* document antérieur, mais publié à la date de dépôt international ou après cette date
- \*L\* document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)
- \*O\* document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens
- \*P\* document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

\*T\* document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention

\*X\* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément

\*Y\* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier

\*G\* document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

9 février 2005

Date d'expédition du présent rapport de recherche internationale

16/02/2005

Nom et adresse postale de l'administration chargée de la recherche internationale  
Office Européen des Brevets, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Fonctionnaire autorisé

Michalski, S

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS		
Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	<p>CHEN DING ET AL: "Centralized content-based Web filtering and blocking: how far can it go?"            SYSTEMS, MAN, AND CYBERNETICS, 1999. IEEE SMC '99 CONFERENCE PROCEEDINGS. 1999 IEEE INTERNATIONAL CONFERENCE ON TOKYO, JAPAN 12-15 OCT. 1999, PISCATAWAY, NJ, USA, IEEE, US, 12 octobre 1999 (1999-10-12), pages 115-119, XP010363557            ISBN: 0-7803-5731-0            page 116, colonne de gauche            page 117, colonne de gauche            page 119, colonne de gauche</p>	1-4, 11-14
X	<p>US 2003/126267 A1 (BRODSKY TOMAS ET AL)            3 juillet 2003 (2003-07-03)</p>	1,3-5,7, 8,10,11, 13-15, 17,18,20
Y	figures 1,4	2,12
Y	alinéas '0002!, '0007!, '0013!, '0015!, '0023!	5,6,15, 16
X	<p>WO 01/98934 A (INFONUTZ LLC ;MORTL WILLIAM M (US)) 27 décembre 2001 (2001-12-27)</p> <p>page 3            page 9            revendication 1</p>	1,3,4, 8-11,13, 14,18-20
Y	<p>FORSYTH D A ET AL: "Identifying nude pictures"            APPLICATIONS OF COMPUTER VISION, 1996. WACV '96., PROCEEDINGS 3RD IEEE WORKSHOP ON SARASOTA, FL, USA 2-4 DEC. 1996, LOS ALAMITOS, CA, USA, IEEE COMPUT. SOC, US, 2 décembre 1996 (1996-12-02), pages 103-108, XP010206418            ISBN: 0-8186-7620-5            abrégé            page 104, colonne de gauche, alinéas 1,2</p>	5,6,15, 16
A	<p>"LOOK AHEAD FILTERING OF INTERNET CONTENT"            IBM TECHNICAL DISCLOSURE BULLETIN, IBM CORP. NEW YORK, US,            vol. 40, no. 12,            1 décembre 1997 (1997-12-01), page 143,            XP000754118            ISSN: 0018-8689</p>	1,11
Y	le document en entier	2,12

# RAPPORT DE RECHERCHE INTERNATIONALE

PCT/EP2004/052571

Document brevet cité au rapport de recherche		Date de publication	Membre(s) de la famille de brevet(s)		Date de publication
US 2003126267	A1	03-07-2003	AU	2002367040 A1	30-07-2003
			WO	03060757 A2	24-07-2003
WO 0198934	A	27-12-2001	AU	6857901 A	02-01-2002
			WO	0198934 A2	27-12-2001